



---

## **Re: Answers to Questions Posed by the Right to Know Advisory Committee/Bulk Records Subcommittee**

### **Introduction**

InforME, as the State's official electronic portal, offers bulk data access as one of its information access services. These services are created from data assembled through value-added processes, from agency repositories for access to authorized InforME customers on behalf of state agencies. The InforME Board is mindful that there are many complex issues related to this discussion and that the services it provides and the terminology it uses varies significantly from others supplying information to the subcommittee. It hopes that this submission will provide information helpful in the committee's deliberations and is committed to supporting the committee's efforts on improving public information access and related issues going forward.

### **1. What is bulk data and how should it be defined?**

#### **Bulk data defined:**

Bulk data is an electronic collection of data composed of information from multiple records, whose primary relationship to each other is their shared origin from a single or multiple databases. The data is generally extracted from a database and provided in a common electronic file type. Bulk data does not include paper copies of records. A request for bulk data is also different from a request for multiple records, if we consider a record to be a "document" such as an individual deed, copy of a license, birth certificate, or even the data collected and stored related to a single event, person or other data element (for example). Requesting multiple records (paper or electronic images) is different from requesting a data file with information from multiple records - the latter is bulk data. Each individual agency determines which bulk data sets are eligible for public request and is responsible for ensuring the state laws, regulations, and privacy requirements are adhered to whenever bulk data is made available for access.

Databases themselves may be proprietary, or complex and unusable to a requester without special software or instructions. A request for data is not the same as a request for a database.

#### **InforME bulk data services defined:**

InforME, as the State's official electronic portal, offers bulk data assembled through value-added processes, from agency repositories (electronic databases) to provide access to authorized InforME customers on behalf of state agencies. InforME may modify the data from its original form in various ways, including removal of sensitive information, selection, sorting or combining records for a particular use or to customer specifications, and reformatting or restructuring of data, as well as a variety of delivery mechanisms including self-service online services and FTP delivery. InforME offers data for one-time access or on recurring schedules such as monthly or quarterly, and manages customer requests, customer service, payments and invoicing.

## **2. What is the appropriate method of determining the cost that a requestor must pay for bulk data?**

Fees for InforME bulk data access are determined by statute, agency rulemaking, or through board oversight as defined in Title 1 §534 5. G. which specifies that fees for InforME services must be:

- Sufficient to maintain, develop, operate and expand InforME on a continuing basis.
- Reasonable but sufficient to support the maximum amount of information and services provided at no charge.
- Sufficient to ensure that, to the extent possible, data custodians do not suffer loss of revenues from sources that are approved or authorized by law due to the operations of InforME.
- Sufficient to ensure that data custodians are reimbursed for the actual costs of providing data to InforME.
- Sufficient to meet the expenses of the statutory InforME board.

InforME's service and data delivery model is designed to enhance electronic access to public information and transactions for citizens and businesses, by pooling fees from all services that have commercial value, to provide those services plus other services which have no commercial value but are needed or desired by Maine citizens and businesses. Recapturing some of the commercial value of the taxpayer-assembled data in order to provide services to Maine citizens and businesses allows InforME to increase the availability of all state electronic information and online services. Revenue from bulk data services supports not only bulk data distributions but also the provisioning of many other electronic services to the public, and also to state agencies. Those other services are then provided to the public and agencies at no cost to them. The elimination of its bulk data services or reductions to pricing would likely reduce InforME's ability to provide access to public information via existing and future portal services, with resulting effect on the ability of state agencies to provide efficient, cost-effective online services for the public.

Examples of InforME public access services supported in part by bulk data services:

- Maine.gov - the State of Maine's official web site
- Hosting of nearly all Maine state agency websites
- Agency webmaster support, tools, and training
- Online Absentee Ballot Request system
- Maine.gov DataShare public data catalog
- HireME online state job application system
- Online Sex Offender Registry Search
- Online State Parks Search
- Unclaimed Property Search/Claim service
- Voter Information service
- Qualifying Contributions for Clean Elections Candidates system
- Maine Organ Donor Registry

### **3. Should a requestor of bulk data be entitled to the records in the format and type of access requested? Should a distinction be made between a requestor seeking access to records and a requestor seeking ownership of records?**

Data distributed via InforME's bulk data services is delivered in an electronic file format that is developed based upon a balance of considerations between minimizing assembly costs and providing the greatest utility to the largest number of customers possible. It is not practical for a requestor to have the right to receive data in any format they desire, as that may be impossible in some cases, and very expensive in others. If multiple formats are already available, then it is reasonable for the requestor to choose which format they prefer.

#### **Access vs. Ownership**

In the digital age, it is very difficult to distinguish sometimes between access and ownership. The information being accessed as bulk data is assembled at taxpayer expense for the purpose of allowing the agency to do its statutorily required work. While agreements can limit rights to use records, it is often difficult to determine, without an after-the-fact audit, whether the records were actually used in accordance with the contract restrictions, once out of sight of InforME or the agency. Similarly, while an agreement can withhold ownership and only grant a license to use, because bulk data is by definition an extract of information from a record, it may also be impossible to tell whether someone's address came from a state record or from other sources. Certainly any information that can *only* come from a state record would be easier to track for compliance purposes. For requestors of information, the issue will be whether they are entitled under "access" to use the data however they wish.

An example may be helpful. Certain requestors use vehicle registration information to send safety recalls on behalf of vehicle manufacturers. Such information is subject to the Driver Privacy Protection Act ("DPPA"), which restricts its use to certain purposes, including safety recalls. If the bulk requester was to assert ownership to the information which it *obtained* from the bulk data -- for example, vehicle owner address information -- and used it for a purpose outside the limits of the DPPA, then this could be a violation of the law. But it might be difficult to establish that the address information came from a state database. However, a vehicle identification number, which is not generally available from other sources, might be more easily proven to have come from a state database.

In the case of vehicle registrations, access might be the appropriate level of permission—access to use for a specific purpose, or any purpose permitted by the DPPA. Ownership is not similarly restricted, so ownership might be inappropriate for such data.

However, in the digital age, again, once a "copy" of the extracted data is delivered to a requestor, it may become very difficult to prevent the requestor from asserting that it purchased non-exclusive "ownership" rather than only a license to access for a specific purpose, and even more difficult to restrict what the owner does with the data in its possession.

### **4. Should the law distinguish between bulk data requests of public records for commercial purposes versus requests for noncommercial purposes?**

Yes. The great majority of InforME bulk data requests are for commercial use, since most individuals have little use for bulk data. Pricing for InforME's services, including bulk data access, to commercial entities is designed to conform with InforME's statutory constraints

(above) and is additionally limited by the data's market value. The provision of bulk data services requires resources for customer service, invoicing, compiling and delivering data, and creating and maintaining electronic systems specifically for the purpose of meeting bulk data requests. By leveraging InforME to provide these services on behalf of all state agencies, state agencies offer an economical value-added service to customers, save significant staff time and expense, and contribute to the support of other InforME services for the public's benefit.

Freedom of access requires provision of access to government records, which obligation is met by making individual records available for viewing. This prevents government processes and decisions from being made secretly or without accountability. Commercial entities have many legitimate reasons for purchasing bulk government data and there is no reason they should not be allowed to do so with the intention of making a profit, so long as those efforts are balanced against the public's privacy interest and do not undermine the government's service delivery missions including public access efforts. Commercial requests for bulk data services are not about "access to government records" as the records are already available in individual formats; rather, they are about the commercial desire for a value-added compilation and delivery of data with commercial value. The market value of bulk data is a cost of business that commercial entities can absorb. Commercial entities should not obtain valuable bulk data for their commercial use at a nominal cost.

On the other side, there are a small number of requests for bulk data for non-commercial purposes such as academic research or news reporting. It makes sense to offer a simpler (less resource intensive) data service for these requestors at lower cost, with use restricted to those limited purposes.

### **How does privacy fit into the discussion?**

InforME is mindful of the threat to privacy for individuals posed by bulk data access that goes beyond personally identifiable information contained in individual records. Wide distribution of large volumes of public information can be subjected to modern computer processing treatments involving multiple databases which can yield a depth of information about individuals that goes beyond the anticipated disclosure. This is generally known as data mining. InforME does not perform data mining and distributes no data, public or not, without authorization by agency data custodians and the InforME Board, and then only in accordance with the requirements of the law.

The difficulty is in permitting bulk record access for worthwhile uses (selective service registration compliance, vehicle safety recalls, voter roll verification, jury duty rolls, and even mortgage marketing) while preventing abuses of bulk data availability.